

## **Multivariate Log-Concave Distributions as a Nearly Parametric Model\***

Dominic Schuhmacher, André Hüsler and Lutz Dümbgen

July 2009 (minor revisions in February 2010 and April 2011)

### **Abstract**

In this paper we show that the family  $\mathcal{P}_d^{(\text{lc})}$  of probability distributions on  $\mathbb{R}^d$  with log-concave densities satisfies a strong continuity condition. In particular, it turns out that weak convergence within this family entails (i) convergence in total variation distance, (ii) convergence of arbitrary moments, and (iii) pointwise convergence of Laplace transforms. In this and several other respects the nonparametric model  $\mathcal{P}_d^{(\text{lc})}$  behaves like a parametric model such as, for instance, the family of all  $d$ -variate Gaussian distributions. As a consequence of the continuity result, we prove the existence of nontrivial confidence sets for the moments of an unknown distribution in  $\mathcal{P}_d^{(\text{lc})}$ . Our results are based on various new inequalities for log-concave distributions which are of independent interest.

**Keywords and phrases.** confidence set, moments, Laplace transform, total variation, weak continuity, weak convergence.

**AMS 2000 subject classification.** 62A01, 62G05, 62G07, 62G15, 62G35

\* Work supported by Swiss National Science Foundation

# 1 Introduction

It is well-known that certain statistical functionals such as moments fail to be weakly continuous on the set of, say, all probability measures on the real line for which these functionals are well-defined. This is the intrinsic reason why it is impossible to construct nontrivial two-sided confidence intervals for such functionals. For the mean and other moments, this fact was pointed out by Bahadur and Savage (1956). Donoho (1988) extended these considerations by noting that some functionals of interest are at least weakly semi-continuous, so that one-sided confidence bounds are possible.

When looking at the proofs of the results just mentioned, one realizes that they often involve rather strange, e.g. multimodal or heavy-tailed, distributions. Natural questions are whether statistical functionals such as moments become weakly continuous and whether honest confidence intervals exist for these functionals if attention is restricted to a suitable nonparametric class of distributions. For instance, one possibility would be to focus on distributions on a given bounded region. But this may be too restrictive or lead to rather conservative procedures.

Alternatively we propose a qualitative constraint. When asking a statistician to draw a typical probability density, she or he will often sketch a bell-shaped, maybe skewed density. This suggests unimodality as a constraint, but this would not rule out heavy tails. In the present paper we favor the stronger though natural constraint of log-concavity, also called strong unimodality. One should note here that additional assumptions such as given bounded support or log-concavity can never be strictly verified based on empirical data alone; see Donoho (1988, Section 2).

Before proceeding with log-concavity, let us consider briefly the parametric model  $\mathcal{N}_d$  of all nondegenerate Gaussian distributions on  $\mathbb{R}^d$ . Suppose that a sequence of distributions  $P_n = N_d(\mu_n, \Sigma_n) \in \mathcal{N}_d$  converges weakly to  $P = N_d(\mu, \Sigma) \in \mathcal{N}_d$ . This is easily shown to be equivalent to  $\mu_n \rightarrow \mu$  and  $\Sigma_n \rightarrow \Sigma$  as  $n \rightarrow \infty$ . But this implies convergence in total variation distance, i.e.

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} |f_n(x) - f(x)| dx = 0,$$

where  $f_n$  and  $f$  denote the Lebesgue densities of  $P_n$  and  $P$ , respectively. Furthermore, weak convergence of  $(P_n)_n$  to  $P$  in  $\mathcal{N}_d$  implies convergence of all moments and pointwise convergence of the Laplace-transforms. That means, for all  $d$ -variate polynomials  $\Pi : \mathbb{R}^d \rightarrow \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \int \Pi(x) f_n(x) dx = \int \Pi(x) f(x) dx,$$

and for arbitrary  $\theta \in \mathbb{R}^d$ ,

$$\lim_{n \rightarrow \infty} \int \exp(\theta^\top x) f_n(x) dx = \int \exp(\theta^\top x) f(x) dx.$$

In the present paper we show that the nonparametric model  $\mathcal{P}_d^{(\text{lc})}$  of all log-concave probability distributions  $P$  on  $\mathbb{R}^d$  has the same properties. Log-concavity of  $P$  means that it admits a Lebesgue density  $f$  of the form

$$f(x) = \exp(\varphi(x))$$

for some concave function  $\varphi : \mathbb{R}^d \rightarrow [-\infty, \infty)$ . Obviously the model  $\mathcal{P}_d^{(\text{lc})}$  contains the parametric family  $\mathcal{N}_d$ . All of its members are unimodal in that the level sets  $\{x \in \mathbb{R}^d : f(x) \geq c\}$ ,  $c > 0$ , are bounded and convex. It is further known that product measures, marginals, convolutions, and weak limits (if a limiting density exists) of log-concave distributions are log-concave; see Dharmadhikari and Joag-dev (1988), Chapter 2. These closedness properties are again shared by the class of Gaussian distributions. The results in the present paper make a substantial contribution to the list of such shared properties and thus promote the view of the model  $\mathcal{P}_d^{(\text{lc})}$  as a viable nonparametric substitute for the Gaussian model  $\mathcal{N}_d$ .

The univariate class  $\mathcal{P}_1^{(\text{lc})}$  has been studied extensively; see Bagnoli and Bergstrom (2005), Dümbgen and Rufibach (2009) and the references therein. Many standard models of univariate distributions belong to this nonparametric family, e.g. all gamma distributions with shape parameter  $\geq 1$ , and all beta distributions with both parameters  $\geq 1$ . Bagnoli and Bergstrom (2005) establish various properties of the corresponding distribution and hazard functions. Nonparametric maximum likelihood estimation of a distribution in  $\mathcal{P}_1^{(\text{lc})}$  has been studied by Pal et al. (2006) and Dümbgen and Rufibach (2009). In particular, the latter two papers provide consistency results for these estimators. The findings of the present paper allow to strengthen these results considerably by showing that consistency in any reasonable sense implies consistency of all moments and, much more generally, consistency of the densities in exponentially weighted total variation distance. Algorithms for the one-dimensional maximum-likelihood estimator are described by Dümbgen et al. (2007) and Dümbgen and Rufibach (2011).

The multivariate class  $\mathcal{P}_d^{(\text{lc})}$  is in various respects more difficult to treat. It has been considered in Dharmadhikari and Joag-dev (1988) and An (1998). Comprehensive treatments of the state of the art in multivariate log-concave density modeling and estimation are Cule et al. (2010) and the survey paper by Walther (2009). An explicit algorithm for the nonparametric maximum likelihood estimator is provided by Cule et al. (2009). Consistency of this estimator has been verified by Cule and Samworth (2010) and Schuhmacher and Dümbgen (2010). Again the results of the present paper allow to transfer consistency properties into much stronger modes of consistency.

The remainder of this paper is organized as follows. In Section 2 we present our main result and some consequences, including an existence proof of non-trivial confidence sets for moments of log-concave distributions. Section 3 collects some basic inequalities for log-concave distributions which are essential for the main results and of independent interest. Most proofs are deferred to Section 4.

## 2 The main results

Let us first introduce some notation. Throughout this paper,  $\|\cdot\|$  stands for Euclidean norm. The closed Euclidean ball with center  $x \in \mathbb{R}^d$  and radius  $\epsilon \geq 0$  is denoted by  $B(x, \epsilon)$ . With  $\text{int}(S)$  and  $\partial S$  we denote the interior and boundary, respectively, of a set  $S \subset \mathbb{R}^d$ .

**Theorem 2.1.** *Let  $P, P_1, P_2, P_3 \dots$  be probability measures in  $\mathcal{P}_d^{(\text{lc})}$  with densities  $f, f_1, f_2, f_3, \dots$ , respectively, such that  $P_n \rightarrow P$  weakly as  $n \rightarrow \infty$ . Then the following two conclusions hold true:*

- (i) *The sequence  $(f_n)$  converges uniformly to  $f$  on any closed set of continuity points of  $f$ .*
- (ii) *Let  $A : \mathbb{R}^d \rightarrow \mathbb{R}$  be a sublinear function, i.e.  $A(x+y) \leq A(x) + A(y)$  and  $A(rx) = rA(x)$  for all  $x, y \in \mathbb{R}^d$  and  $r \geq 0$ . If*

$$f(x) \exp(A(x)) \rightarrow 0 \quad \text{as } \|x\| \rightarrow \infty, \quad (2.1)$$

*then  $\int_{\mathbb{R}^d} \exp(A(x)) f(x) dx < \infty$  and*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \exp(A(x)) |f_n(x) - f(x)| dx = 0. \quad (2.2)$$

It is well-known from convex analysis that  $\varphi = \log f$  is continuous on  $\text{int}(\{\varphi > -\infty\}) = \text{int}(\{f > 0\})$ . Hence the discontinuity points of  $f$ , if any, are contained in  $\partial\{f > 0\}$ . But  $\{f > 0\}$  is a convex set, so its boundary has Lebesgue measure zero (cf. Lang 1986). Therefore Part (i) of Theorem 2.1 implies that  $(f_n)_n$  converges to  $f$  pointwise almost everywhere.

Note also that  $f(x) \leq C_1 \exp(-C_2 \|x\|)$  for suitable constants  $C_1 = C_1(f) > 0$  and  $C_2 = C_2(f) > 0$ ; see Corollary 3.4 in Section 3. Hence one may take  $A(x) = c\|x\|$  for any  $c \in [0, C_2]$  in order to satisfy (2.1). Theorem 2.1 is a multivariate version of Hüsler (2008, Theorem 2.1). It is also more general than findings of Cule and Samworth (2010) who treated the special case of  $A(x) = \epsilon\|x\|$  for some small  $\epsilon > 0$  with different techniques.

Before presenting the conclusions about moments and moment generating functions announced in the introduction, let us provide some information about the moment generating functions of distributions in  $\mathcal{P}_d^{(\text{lc})}$ :

**Proposition 2.2.** *For a distribution  $P \in \mathcal{P}_d^{(\text{lc})}$  let  $\Theta(P)$  be the set of all  $\theta \in \mathbb{R}^d$  such that  $\int \exp(\theta^\top x) P(dx) < \infty$ . This set  $\Theta(P)$  is convex, open and contains 0. Let  $\theta \in \mathbb{R}^d$  and  $\epsilon > 0$  such that  $B(\theta, \epsilon) \subset \Theta(P)$ . Then*

$$A(x) := \theta^\top x + \epsilon\|x\|$$

*defines a sublinear function  $A$  on  $\mathbb{R}^d$  such that the density  $f$  of  $P$  satisfies*

$$\lim_{\|x\| \rightarrow \infty} \exp(A(x)) f(x) = 0.$$

Note that for any  $d$ -variate polynomial  $\Pi$  and arbitrary  $\epsilon > 0$  there exists an  $R = R(\Pi, \epsilon) > 0$  such that  $|\Pi(x)| \leq \exp(\epsilon\|x\|)$  for  $\|x\| > R$ . Hence part (ii) of Theorem 2.1 and Proposition 2.2 entail the first part of the following theorem:

**Theorem 2.3.** *Under the conditions of Theorem 2.1, for any  $\theta \in \Theta(P)$  and arbitrary  $d$ -variate polynomials  $\Pi : \mathbb{R}^d \rightarrow \mathbb{R}$ , the integral  $\int_{\mathbb{R}^d} \exp(\theta^\top x) |\Pi(x)| f(x) dx$  is finite and*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \exp(\theta^\top x) |\Pi(x)| |f_n(x) - f(x)| dx = 0.$$

Moreover, for any  $\theta \in \mathbb{R}^d \setminus \Theta(P)$ ,

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \exp(\theta^\top x) f_n(x) dx = \infty.$$

**Existence of nontrivial confidence sets for moments.** With the previous results we can prove the existence of confidence sets for arbitrary moments, modifying Donoho's (1988) recipe. Let  $\mathcal{H} = \mathcal{H}_d$  denote the set of all closed halfspaces in  $\mathbb{R}^d$ . For two probability measures  $P$  and  $Q$  on  $\mathbb{R}^d$  let

$$\|P - Q\|_{\mathcal{H}} := \sup_{H \in \mathcal{H}} |P(H) - Q(H)|.$$

It is well-known from empirical process theory (e.g. van der Vaart and Wellner 1996, Section 2.19) that for any  $\alpha \in (0, 1)$  there exists a universal constant  $c_{\alpha, d}$  such that

$$\mathbb{P}\left(\|\hat{P}_n - P\|_{\mathcal{H}} \geq n^{-1/2} c_{\alpha, d}\right) \leq \alpha$$

for arbitrary distributions  $P$  on  $\mathbb{R}^d$  and the empirical distribution  $\hat{P}_n$  of independent random vectors  $X_1, X_2, \dots, X_n \sim P$ . In particular, Massart's (1990) inequality yields the constant  $c_{\alpha, 1} = (\log(2/\alpha)/2)^{1/2}$ .

Under the assumption that  $P \in \mathcal{P}_d^{(\text{lc})}$ , a  $(1 - \alpha)$ -confidence set for the distribution  $P$  is given by

$$C_{\alpha, n} = C_{\alpha, n}(X_1, X_2, \dots, X_n) := \left\{ Q \in \mathcal{P}_d^{(\text{lc})} : \|Q - \hat{P}_n\|_{\mathcal{H}} \leq n^{-1/2} c_{\alpha, d} \right\}.$$

This entails simultaneous  $(1 - \alpha)$ -confidence sets for all integrals  $\int \Pi(x) P(dx)$ , where  $\Pi : \mathbb{R}^d \rightarrow \mathbb{R}$  is an arbitrary polynomial, namely,

$$C_{\alpha, n}^{(\Pi)} = C_{\alpha, n}^{(\Pi)}(X_1, X_2, \dots, X_n) := \left\{ \int \Pi(x) Q(dx) : Q \in C_{\alpha, n} \right\}.$$

Since convergence with respect to  $\|\cdot\|_{\mathcal{H}}$  implies weak convergence, Theorem 2.3 implies the consistency of the confidence sets  $C_{\alpha, n}^{(\Pi)}$ , in the sense that

$$\sup_{t \in C_{\alpha, n}^{(\Pi)}} \left| t - \int \Pi(x) P(dx) \right| \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty.$$

Note that this construction proves existence of honest simultaneous confidence sets for arbitrary moments. But their explicit computation requires substantial additional work and is beyond the scope of the present paper.

### 3 Various inequalities for $\mathcal{P}_d^{(\text{lc})}$

In this section we provide a few inequalities for log-concave distributions which are essential for the main result or are of independent interest. Let us first introduce some notation. The convex hull of a nonvoid set  $S \subset \mathbb{R}^d$  is denoted by  $\text{conv}(S)$ , the Lebesgue measure of a Borel set  $S \subset \mathbb{R}^d$  by  $|S|$ .

#### 3.1 Inequalities for general dimension

**Lemma 3.1.** *Let  $P \in \mathcal{P}_d^{(\text{lc})}$  with density  $f$ . Let  $x_0, x_1, \dots, x_d$  be fixed points in  $\mathbb{R}^d$  such that  $\Delta := \text{conv}\{x_0, x_1, \dots, x_d\}$  has nonvoid interior. Then*

$$\prod_{j=0}^d f(x_j) \leq \left( \frac{P(\Delta)}{|\Delta|} \right)^{d+1}.$$

*Suppose that  $x_1, x_2, \dots, x_d \in \{f > 0\}$ , and define  $\tilde{f}(x_1, \dots, x_d) := \left( \prod_{i=1}^d f(x_i) \right)^{1/d}$ . Then*

$$\frac{f(x_0)}{\tilde{f}(x_1, \dots, x_d)} \leq \left( \frac{P(\Delta)}{\tilde{f}(x_1, \dots, x_d) |\Delta|} \right)^{d+1}$$

*If the right hand side is less than or equal to one, then*

$$\frac{f(x_0)}{\tilde{f}(x_1, \dots, x_d)} \leq \exp\left(d - d \frac{\tilde{f}(x_1, \dots, x_d) |\Delta|}{P(\Delta)}\right).$$

This lemma entails various upper bounds including a subexponential tail bound for log-concave densities.

**Lemma 3.2.** *Let  $x_0, x_1, \dots, x_d \in \mathbb{R}^d$  and  $\Delta$  as in Lemma 3.1. Then for any  $P \in \mathcal{P}_d^{(\text{lc})}$  with density  $f$  such that  $x_0, x_1, \dots, x_d \in \{f > 0\}$  and arbitrary  $y \in \Delta$ ,*

$$\min_{i=0, \dots, d} f(x_i) \leq f(y) \leq \left( \frac{P(\Delta)}{|\Delta|} \right)^{d+1} \left( \min_{i=0, \dots, d} f(x_i) \right)^{-d}.$$

**Lemma 3.3.** *Let  $x_0, x_1, \dots, x_d \in \mathbb{R}^d$  as in Lemma 3.1. Then there exists a constant  $C = C(x_0, x_1, \dots, x_d) > 0$  with the following property: For any  $P \in \mathcal{P}_d^{(\text{lc})}$  with density  $f$  such that  $x_0, x_1, \dots, x_d \in \{f > 0\}$  and arbitrary  $y \in \mathbb{R}^d$ ,*

$$f(y) \leq \max_{i=0, \dots, d} f(x_i) H\left(C \min_{i=0, \dots, d} f(x_i) (1 + \|y\|^2)^{1/2}\right),$$

where

$$H(t) := \begin{cases} t^{-(d+1)} & \text{for } t \in [0, 1], \\ \exp(d - dt) & \text{for } t \geq 1. \end{cases}$$

**Corollary 3.4.** *For any  $P \in \mathcal{P}_d^{(\text{lc})}$  with density  $f$  there exist constants  $C_1 = C_1(P) > 0$  and  $C_2 = C_2(P) > 0$  such that*

$$f(x) \leq C_1 \exp(-C_2 \|x\|) \quad \text{for all } x \in \mathbb{R}^d.$$

## 3.2 Inequalities for dimension one

In the special case  $d = 1$  we denote the cumulative distribution function of  $P$  with  $F$ . The hazard functions  $f/F$  and  $f/(1 - F)$  have the following properties:

**Lemma 3.5.** *The function  $f/F$  is non-increasing on  $\{x : 0 < F(x) \leq 1\}$ , and the function  $f/(1 - F)$  is non-decreasing on  $\{x : 0 \leq F(x) < 1\}$ .*

*Let  $t_\ell := \inf\{f > 0\}$  and  $t_u := \sup\{f > 0\}$ . Then*

$$\begin{aligned} \lim_{t \downarrow t_\ell} \frac{f(t)}{F(t)} &= \infty \quad \text{if } t_\ell > -\infty, \\ \lim_{t \uparrow t_u} \frac{f(t)}{1 - F(t)} &= \infty \quad \text{if } t_u < \infty. \end{aligned}$$

The monotonicity properties of the hazard functions  $f/F$  and  $f/(1 - F)$  have been noted by An (1998) and Bagnoli and Bergstrom (2005). For the reader's convenience a complete proof of Lemma 3.5 will be given.

The next lemma provides an inequality for  $f$  in terms of its first and second moments:

**Lemma 3.6.** *Let  $\mu$  and  $\sigma$  be the mean and standard deviation, respectively, of the distribution  $P$ . Then for arbitrary  $x_o \in \mathbb{R}$ ,*

$$f(x_o)^2 \leq \frac{2F(x_o)^3 + 2(1 - F(x_o))^3}{(x_o - \mu)^2 + \sigma^2}.$$

*Equality holds if, and only if,  $f$  is log-linear on both  $(-\infty, x_o]$  and  $[x_o, \infty)$ .*

## 4 Proofs

### 4.1 Proofs for Section 3

Our proof of Lemma 3.1 is based on a particular representation of Lebesgue measure on simplices: Let

$$\Delta_o := \left\{ u \in [0, 1]^d : \sum_{i=1}^d u_i \leq 1 \right\}.$$

Then for any measurable function  $h : \Delta_o \rightarrow [0, \infty)$ ,

$$\int_{\Delta_o} h(u) du = \frac{1}{d!} \mathbb{E} h(B_1, B_2, \dots, B_d),$$

where  $B_i := E_i / \sum_{j=0}^d E_j$  with independent, standard exponentially distributed random variables  $E_0, E_1, \dots, E_d$ . This follows from general considerations about gamma and multivariate beta distributions, e.g. in Cule and D mbgen (2008). In particular,  $|\Delta_o| = 1/d!$ . Moreover, each variable  $B_i$  is beta distributed with parameters 1 and  $d$ , and  $\mathbb{E}(B_i) = 1/(d+1)$ .

**Proof of Lemma 3.1.** Any point  $x \in \Delta$  may be written as

$$x(u) := x_0 + \sum_{i=1}^d u_i (x_i - x_0) = \sum_{i=0}^d u_i x_i$$

for some  $u \in \Delta_o$ , where  $u_0 := 1 - \sum_{i=1}^d u_i$ . In particular,

$$\frac{|\Delta|}{|\Delta_o|} = |\det(x_1 - x_0, x_2 - x_0, \dots, x_d - x_0)|.$$

By concavity of  $\varphi := \log f$ ,

$$\varphi(x(u)) \geq \sum_{i=0}^d u_i \varphi(x_i)$$

for any  $u = (u_i)_{i=1}^d \in \Delta_o$  and  $u_0 = 1 - \sum_{i=1}^d u_i$ . Hence

$$\frac{P(\Delta)}{|\Delta|} = \frac{1}{|\Delta_o|} \int_{\Delta_o} \exp(\varphi(x(u))) du = \mathbb{E} \exp\left(\varphi\left(\sum_{i=0}^d B_i x_i\right)\right) \geq \mathbb{E} \exp\left(\sum_{i=0}^d B_i \varphi(x_i)\right),$$

and by Jensen's inequality, the latter expected value is not less than

$$\exp\left(\sum_{i=0}^d \mathbb{E}(B_i) \varphi(x_i)\right) = \exp\left(\frac{1}{d+1} \sum_{i=0}^d \varphi(x_i)\right) = \left(\prod_{i=0}^d f(x_i)\right)^{1/(d+1)}.$$

This yields the first assertion of the lemma.

The inequality  $\prod_{i=0}^d f(x_i) \leq (P(\Delta)/|\Delta|)^{d+1}$  may be rewritten as

$$f(x_0) \tilde{f}(x_1, \dots, x_d)^d \leq \left(\frac{P(\Delta)}{|\Delta|}\right)^{d+1},$$

and dividing both sides by  $\tilde{f}(x_1, \dots, x_d)^{d+1}$  yields the second assertion.

As to the third inequality, suppose that  $f(x_0) \leq \tilde{f}(x_1, \dots, x_d)$ , which is equivalent to  $\varphi_0 := \varphi(x_0)$  being less than or equal to  $\bar{\varphi} := \log \tilde{f}(x_1, \dots, x_d) = d^{-1} \sum_{i=1}^d \varphi(x_i)$ . Then

$$\frac{P(\Delta)}{|\Delta|} \geq \mathbb{E} \exp\left(\sum_{i=0}^d B_i \varphi(x_i)\right) = \mathbb{E} \exp\left(B_0 \varphi_0 + (1 - B_0) \sum_{i=1}^d \tilde{B}_i \varphi(x_i)\right),$$



where  $\tilde{B}_i := E_i / \sum_{j=1}^d E_j$  for  $1 \leq i \leq d$ . It is well-known (e.g. Cule and Dümmbgen 2008) that  $B_0$  and  $(\tilde{B}_i)_{i=1}^d$  are stochastically independent, where  $\mathbb{E}(\tilde{B}_i) = 1/d$ . Hence it follows from Jensen's inequality and  $B_0 \sim \text{Beta}(1, d)$  that

$$\begin{aligned}
\frac{P(\Delta)}{|\Delta|} &\geq \mathbb{E} \mathbb{E} \left( \exp \left( B_0 \varphi_0 + (1 - B_0) \sum_{i=1}^d \tilde{B}_i \varphi(x_i) \right) \middle| B_0 \right) \\
&\geq \mathbb{E} \exp \left( \mathbb{E} \left( B_0 \varphi_0 + (1 - B_0) \sum_{i=1}^d \tilde{B}_i \varphi(x_i) \middle| B_0 \right) \right) \\
&= \mathbb{E} \exp(B_0 \varphi_0 + (1 - B_0) \bar{\varphi}) \\
&= \int_0^1 d(1-t)^{d-1} \exp(t \varphi_0 + (1-t) \bar{\varphi}) dt \\
&= \tilde{f}(x_1, \dots, x_d) \int_0^1 d(1-t)^{d-1} \exp(-t(\bar{\varphi} - \varphi_0)) dt \\
&\geq \tilde{f}(x_1, \dots, x_d) \int_0^1 d(1-t)^{d-1} \exp(\log(1-t)(\bar{\varphi} - \varphi_0)) dt \\
&= \tilde{f}(x_1, \dots, x_d) \int_0^1 d(1-t)^{\bar{\varphi} - \varphi_0 + d - 1} dt \\
&= \tilde{f}(x_1, \dots, x_d) \frac{d}{d + \bar{\varphi} - \varphi_0}.
\end{aligned}$$

Thus  $\bar{\varphi} - \varphi_0 \geq d \tilde{f}(x_1, \dots, x_d) |\Delta| / P(\Delta) - d$ , which is equivalent to

$$\frac{f(x_0)}{\tilde{f}(x_1, \dots, x_d)} \leq \exp \left( d - d \frac{\tilde{f}(x_1, \dots, x_d) |\Delta|}{P(\Delta)} \right). \quad \square$$

We first prove Lemma 3.3 because this provides a tool for the proof of Lemma 3.2 as well.

**Proof of Lemma 3.3.** At first we investigate how the size of  $\Delta$  changes if we replace one of its vertices with another point. Note that for any fixed index  $j \in \{0, 1, \dots, d\}$ ,

$$|\det(x_i - x_j : i \neq j)| = |\det(X)| \quad \text{with} \quad X := \begin{pmatrix} x_0 & x_1 & \dots & x_d \\ 1 & 1 & \dots & 1 \end{pmatrix}.$$

Moreover, any point  $y \in \mathbb{R}^d$  has a unique representation  $y = \sum_{i=0}^d \lambda_i x_i$  with scalars  $\lambda_0, \lambda_1, \dots, \lambda_d$  summing to one. Namely,

$$(\lambda_i)_{i=0}^d = X^{-1} \begin{pmatrix} y \\ 1 \end{pmatrix}.$$

Hence the set  $\Delta_j(y) := \text{conv}(\{x_i : i \neq j\} \cup \{y\})$  has Lebesgue measure

$$\begin{aligned}
|\Delta_j(y)| &= \frac{1}{d!} \left| \det \begin{pmatrix} x_0 & \dots & x_{j-1} & y & x_{j+1} & \dots & x_d \\ 1 & \dots & 1 & 1 & 1 & \dots & 1 \end{pmatrix} \right| \\
&= \frac{1}{d!} \left| \sum_{i=0}^d \lambda_i \det \begin{pmatrix} x_0 & \dots & x_{j-1} & x_i & x_{j+1} & \dots & x_d \\ 1 & \dots & 1 & 1 & 1 & \dots & 1 \end{pmatrix} \right| \\
&= \frac{1}{d!} |\lambda_j| |\det(X)| \\
&= |\lambda_j| |\Delta|.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\max_{j=0,1,\dots,d} |\Delta_j(y)| &= |\Delta| \max_{j=0,1,\dots,d} |\lambda_j| \\
&= |\Delta| \left\| X^{-1} \begin{pmatrix} y \\ 1 \end{pmatrix} \right\|_{\infty} \\
&\geq |\Delta| (d+1)^{-1/2} \left\| X^{-1} \begin{pmatrix} y \\ 1 \end{pmatrix} \right\| \\
&\geq |\Delta| (d+1)^{-1/2} \sigma_{\max}(X)^{-1} (\|y\|^2 + 1)^{1/2},
\end{aligned}$$

where  $\sigma_{\max}(X) > 0$  is the largest singular value of  $X$ .

Now we consider any log-concave probability density  $f$ . Let  $f_{\min}$  and  $f_{\max}$  denote the minimum and maximum, respectively, of  $\{f(x_i) : i = 0, \dots, d\}$ , where  $f_{\min}$  is assumed to be greater than zero. Applying Lemma 3.1 to  $\Delta_j(y)$  in place of  $\Delta$  with suitably chosen index  $j$ , we may conclude that

$$f(y) \leq f_{\max} (C f_{\min} (\|y\|^2 + 1)^{1/2})^{-(d+1)},$$

where  $C = C(x_0, \dots, x_d) := |\Delta| (d+1)^{-1/2} \sigma_{\max}(X)^{-1}$ . Moreover, in case of  $C f_{\min} (\|y\|^2 + 1)^{1/2} \geq 1$ ,

$$f(y) \leq f_{\max} \exp(d - dC f_{\min} (\|y\|^2 + 1)^{1/2}). \quad \square$$

**Proof of Lemma 3.2.** Let  $y \in \Delta$ , i.e.  $y = \sum_{i=0}^d \lambda_i x_i$  with a unique vector  $\lambda = (\lambda_i)_{i=0}^d$  in  $[0, 1]^{d+1}$  whose components sum to one. With  $\Delta_j(y)$  as in the proof of Lemma 3.3, elementary calculations reveal that

$$\Delta = \bigcup_{j \in J} \Delta_j(y),$$

where  $J := \{j : \lambda_j > 0\}$ . Moreover, all these simplices  $\Delta_j(y)$ ,  $j \in J$ , have nonvoid interior, and  $|\Delta_j(y) \cap \Delta_k(y)| = 0$  for different  $j, k \in J$ . Consequently it follows from Lemma 3.1 that

$$\begin{aligned} \frac{P(\Delta)}{|\Delta|} &= \sum_{j \in J} \frac{|\Delta_j(y)|}{|\Delta|} \cdot \frac{P(\Delta_j(y))}{|\Delta_j(y)|} \\ &\geq \sum_{j \in J} \frac{|\Delta_j(y)|}{|\Delta|} \cdot \left( f(y) \prod_{i \neq j} f(x_i) \right)^{1/(d+1)} \\ &\geq \sum_{j \in J} \frac{|\Delta_j(y)|}{|\Delta|} \cdot f(y)^{1/(d+1)} \left( \min_{i=0, \dots, d} f(x_i) \right)^{d/(d+1)} \\ &= f(y)^{1/(d+1)} \left( \min_{i=0, \dots, d} f(x_i) \right)^{d/(d+1)}. \end{aligned}$$

This entails the asserted upper bound for  $f(y)$ . The lower bound follows from the elementary fact that any concave function on the simplex  $\Delta$  attains its minimal value in one of the vertices  $x_0, x_1, \dots, x_d$ .  $\square$

**Proof of Lemma 3.5.** We only prove the assertions about  $f/(1 - F)$ . Considering the distribution function  $\tilde{F}(x) := 1 - F(-x)$  with log-concave density  $\tilde{f}(x) = f(-x)$  then yields the corresponding properties of  $f/F$ .

Note that  $\{F < 1\} = (-\infty, t_u)$ . On  $\{f = 0\} \cap (-\infty, t_u)$ , the function  $f/(1 - F)$  is equal to zero. For  $t \in \{f > 0\} \cap (-\infty, t_u)$ ,

$$\frac{f(t)}{1 - F(t)} = \left( \int_0^\infty \exp(\varphi(t+x) - \varphi(t)) dx \right)^{-1}$$

is non-decreasing in  $t$ , because  $t \mapsto \varphi(t+x) - \varphi(t)$  is non-increasing in  $t \in \{f > 0\}$  for any fixed  $x > 0$ , due to concavity of  $\varphi$ .

In case of  $t_u < \infty$ , fix any point  $s \in (t_\ell, t_u)$ . Then for  $s \leq t < t_u$ ,

$$\begin{aligned} \frac{f(t)}{1 - F(t)} &= \left( \int_t^{t_u} \exp(\varphi(x) - \varphi(t)) dx \right)^{-1} \\ &\geq \left( \int_t^{t_u} \exp(\varphi'(s+)(x-t)) dx \right)^{-1} \\ &\geq \left( \exp(\min(\varphi'(s+), 0)(t_u - t))(t_u - t) \right)^{-1} \\ &\rightarrow \infty \quad \text{as } t \uparrow t_u. \end{aligned}$$

$\square$

**Proof of Lemma 3.6.** The asserted upper bound for  $f(t_o)$  is strictly positive and continuous in  $t_o$ . Hence it suffices to consider a point  $t_o$  with  $0 < F(t_o) < 1$ . Since  $(x_o - \mu)^2 + \sigma^2$

equals  $\int (x - x_o)^2 f(x) dx$ , we try to bound the latter integral from above. To this end, let  $g$  be a piecewise loglinear probability density, namely,

$$g(x) := \begin{cases} f(x_o) \exp(-a|x - x_o|) & \text{if } x \leq x_o, \\ f(x_o) \exp(-b|x - x_o|) & \text{if } x \geq x_o, \end{cases}$$

with  $a := f(x_o)/F(x_o)$  and  $b := f(x_o)/(1 - F(x_o))$ , so that

$$\int_{-\infty}^{x_o} (g - f)(x) dx = \int_{x_o}^{\infty} (g - f)(x) dx = 0.$$

By concavity of  $\log f$ , there are real numbers  $r < x_o < s$  such that  $f \geq g$  on  $(r, s)$  and  $f \leq g$  on  $\mathbb{R} \setminus [r, s]$ . Consequently,

$$\begin{aligned} \int (x - x_o)^2 (f - g)(x) dx &= \int_{-\infty}^{x_o} \underbrace{[(x - x_o)^2 - (r - x_o)^2]}_{\leq 0} (f - g)(x) dx \\ &\quad + \int_{x_o}^{\infty} \underbrace{[(x - x_o)^2 - (s - x_o)^2]}_{\leq 0} (f - g)(x) dx \\ &\leq 0, \end{aligned}$$

with equality if, and only if,  $f = g$ . Now the assertion follows from

$$\begin{aligned} \int (x - x_o)^2 g(x) dx &= f(x_o) \left( \int_0^{\infty} t^2 \exp(-at) dt + \int_0^{\infty} t^2 \exp(-bt) dt \right) \\ &= \frac{2F(x_o)^3 + 2(1 - F(x_o))^3}{f(x_o)^2}. \end{aligned} \quad \square$$

## 4.2 Proof of the main results

Note first that  $\{f > 0\}$  is a convex set with nonvoid interior. For notational convenience we may and will assume that

$$0 \in \text{int}\{f > 0\}.$$

For if  $x_o$  is any fixed interior point of  $\{f > 0\}$  we could just shift the coordinate system and consider the densities  $\tilde{f} := f(x_o + \cdot)$  and  $\tilde{f}_n := f_n(x_o + \cdot)$  in place of  $f$  and  $f_n$ , respectively. Note also that  $A(x_o + x) - A(x) \in [-A(-x_o), A(x_o)]$ , due to subadditivity of  $A$ .

In our proof of Theorem 2.1, Part (i), we utilize two simple inequalities for log-concave densities:

**Lemma 4.1.** *Let  $x_0, x_1, \dots, x_d \in \mathbb{R}^d$  such that  $\Delta := \text{conv}\{x_0, x_1, \dots, x_d\}$  has nonvoid interior. For  $j = 0, 1, \dots, d$  define the “corner simplex”*

$$\Delta_j := \{2x_j - x : x \in \Delta\},$$

i.e. the reflection of  $\Delta$  at the point  $x_j$ . Let  $P \in \mathcal{P}_d^{(\text{lc})}$  with density  $f = \exp \circ \varphi$ . If  $P(\Delta_j) > 0$  for all  $j = 0, 1, \dots, d$ , then  $\Delta \subset \text{int}\{f > 0\}$ , and

$$\begin{aligned} \min_{j=0,1,\dots,d} \log \frac{P(\Delta_j)}{|\Delta|} &\leq \min_{x \in \Delta} \varphi(x) \leq \log \frac{P(\Delta)}{|\Delta|} \\ &\leq \max_{x \in \Delta} \varphi(x) \leq (d+1) \log \frac{P(\Delta)}{|\Delta|} - d \min_{j=0,1,\dots,d} \log \frac{P(\Delta_j)}{|\Delta|}. \end{aligned}$$

Figure 4.1 illustrates the definition of the corner simplices and a key statement in the proof of Lemma 4.1.

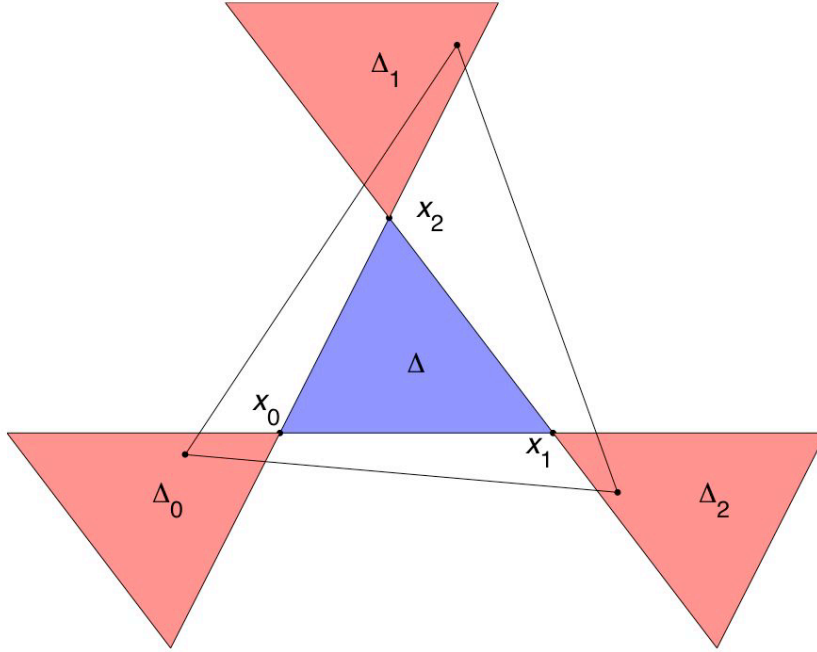


Figure 4.1: A simplex  $\Delta$  and its corner simplices  $\Delta_j$ .

**Lemma 4.2.** Suppose that  $B(0, \delta) \subset \{f > 0\}$  for some  $\delta > 0$ . For  $t \in (0, 1)$  define  $\delta_t := (1 - t)\delta / (1 + t)$ . Then for any  $y \in \mathbb{R}^d$ ,

$$\sup_{x \in B(y, \delta_t)} f(x) \leq \left( \inf_{v \in B(0, \delta)} f(v) \right)^{1-1/t} \left( \frac{P(B(ty, \delta_t))}{|B(ty, \delta_t)|} \right)^{1/t}.$$

This lemma involves three closed balls  $B(0, \delta)$ ,  $B(ty, \delta_t)$  and  $B(y, \delta_t)$ ; see Figure 4.2 for an illustration of these and the key argument of the proof.

**Proof of Lemma 4.1.** Suppose that all corner simplices satisfy  $P(\Delta_j) > 0$ . Then for  $j = 0, 1, \dots, d$  there exists an interior point  $z_j$  of  $\Delta_j$  with  $f(z_j) > 0$ , that means,  $z_j = 2x_j -$

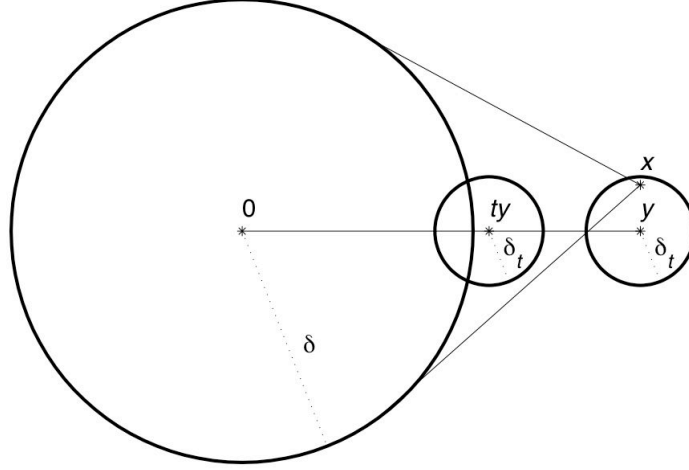


Figure 4.2: The three closed balls in Lemma 4.2.

$\sum_{i=0}^d \lambda_{ij} x_i$  with positive numbers  $\lambda_{ij}$  such that  $\sum_{i=0}^d \lambda_{ij} = 1$ . With the matrices

$$X := \begin{pmatrix} x_0 & x_1 & \dots & x_d \\ 1 & 1 & \dots & 1 \end{pmatrix}, \quad Z := \begin{pmatrix} z_0 & z_1 & \dots & z_d \\ 1 & 1 & \dots & 1 \end{pmatrix} \quad \text{and} \quad \Lambda := \begin{pmatrix} \lambda_{00} & \dots & \lambda_{0d} \\ \vdots & & \vdots \\ \lambda_{d0} & \dots & \lambda_{dd} \end{pmatrix}$$

in  $\mathbb{R}^{(d+1) \times (d+1)}$  we may write

$$Z = X(2I - \Lambda).$$

But the matrix  $2I - \Lambda$  is nonsingular with inverse

$$M := (2I - \Lambda)^{-1} = 2^{-1}(I - 2^{-1}\Lambda)^{-1} = \sum_{\ell=0}^{\infty} 2^{-(\ell+1)} \Lambda^{\ell}.$$

The latter power series converges, because  $\Lambda^{\ell}$  has positive components for all  $\ell \geq 1$ , and via induction on  $\ell \geq 0$  one can show that all columns of  $\Lambda^{\ell}$  sum to one. Consequently,  $X = ZM$ , i.e. for each index  $j$ , the point  $x_j$  may be written as  $\sum_{i=0}^d \mu_{ij} z_i$  with positive numbers  $\mu_{ij}$  such that  $\sum_{i=0}^d \mu_{ij} = 1$ . This entails that  $\Delta$  is a subset of  $\text{int conv}\{z_0, z_1, \dots, z_d\} \subset \text{int}\{f > 0\}$ ; see also Figure 4.1.

Since  $\min_{x \in \Delta} f(x) \leq P(\Delta)/|\Delta| \leq \max_{x \in \Delta} f(x)$ , the inequalities

$$\min_{x \in \Delta} \varphi(x) \leq \log \frac{P(\Delta)}{|\Delta|} \leq \max_{x \in \Delta} \varphi(x)$$

are obvious. By concavity of  $\varphi$ , its minimum over  $\Delta$  equals  $\varphi(x_{j_o})$  for some index  $j_o \in \{0, 1, \dots, d\}$ . But then for arbitrary  $x \in \Delta$  and  $y := 2x_{j_o} - x \in \Delta_{j_o}$ , it follows from  $x_{j_o} = 2^{-1}(x + y)$  and concavity of  $\varphi$  that

$$\varphi(x_{j_o}) \geq \frac{\varphi(x) + \varphi(y)}{2} \geq \frac{\varphi(x_{j_o}) + \varphi(y)}{2},$$

so that  $\varphi \leq \varphi(x_{j_o})$  on  $\Delta_{j_o}$ . Hence

$$\min_{x \in \Delta} \varphi(x) = \varphi(x_{j_o}) \geq \log \frac{P(\Delta_{j_o})}{|\Delta|}.$$

Finally, Lemma 3.2 entails that

$$\begin{aligned} \max_{x \in \Delta} \varphi(x) &\leq (d+1) \log \frac{P(\Delta)}{|\Delta|} - d \min_{j=0,1,\dots,d} \varphi(x_j) \\ &\leq (d+1) \log \frac{P(\Delta)}{|\Delta|} - d \min_{j=0,1,\dots,d} \log \frac{P(\Delta_j)}{|\Delta|}. \end{aligned} \quad \square$$

**Proof of Lemma 4.2.** The main point is to show that for any point  $x \in B(y, \delta_t)$ ,

$$B(ty, \delta_t) \subset (1-t)B(0, \delta) + tx,$$

i.e. any point  $w \in B(ty, \delta_t)$  may be written as  $(1-t)v + tx$  for a suitable  $v \in B(0, \delta)$ ; see also Figure 4.2. But note that the equation  $(1-t)v + tx = w$  is equivalent to  $v = (1-t)^{-1}(w - tx)$ . This vector  $v$  belongs indeed to  $B(0, \delta)$ , because

$$\|v\| = (1-t)^{-1}\|w - tx\| = (1-t)^{-1}\|w - ty + t(y - x)\| \leq (1-t)^{-1}(\delta_t + t\delta_t) = \delta$$

by definition of  $\delta_t$ .

This consideration shows that for any point  $x \in B(y, \delta_t)$  and any point  $w \in B(ty, \delta_t)$ ,

$$f(w) \geq f(v)^{1-t} f(x)^t \geq J_0^{1-t} f(x)^t$$

with  $v = (1-t)^{-1}(w - tx) \in B(0, \delta)$  and  $J_0 := \inf_{v \in B(0, \delta)} f(v)$ . Averaging this inequality with respect to  $w \in B(ty, \delta_t)$  yields

$$\frac{P(B(ty, \delta_t))}{|B(ty, \delta_t)|} \geq J_0^{1-t} f(x)^t.$$

Since  $x \in B(y, \delta_t)$  is arbitrary, this entails the assertion of Lemma 4.2.  $\square$

**Proof of Theorem 2.1, Part (i).** Our proof is split into three steps.

**Step 1:** The sequence  $(f_n)_n$  converges to  $f$  uniformly on any compact subset of  $\text{int}\{f > 0\}$ .

By compactness, this claim is a consequence of the following statement: For any interior point  $y$  of  $\{f > 0\}$  and any  $\eta > 0$  there exists a neighborhood  $\Delta(y, \eta)$  of  $y$  such that

$$\limsup_{n \rightarrow \infty} \sup_{x \in \Delta(y, \eta)} \left| \frac{f_n(x)}{f(x)} - 1 \right| \leq \eta.$$

To prove the latter statement, fix any number  $\epsilon \in (0, 1)$ . Since  $f$  is continuous on  $\text{int}\{f > 0\}$ , there exists a simplex  $\Delta = \text{conv}\{x_0, x_1, \dots, x_d\}$  such that  $y \in \text{int } \Delta$  and

$$f \in [(1 - \epsilon)f(y), (1 + \epsilon)f(y)] \quad \text{on} \quad \Delta \cup \Delta_0 \cup \Delta_1 \cup \dots \cup \Delta_d$$

with the corner simplices  $\Delta_j$  defined as in Lemma 4.1. Since the boundary of any simplex  $\tilde{\Delta}$  is contained in the union of  $d + 1$  hyperplanes, it satisfies  $P(\partial\tilde{\Delta}) = 0$ , so that weak convergence of  $(P_n)_n$  to  $P$  implies that

$$\lim_{n \rightarrow \infty} P_n(\tilde{\Delta}) = P(\tilde{\Delta}).$$

Therefore it follows from Lemma 4.1 that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \inf_{x \in \Delta} \frac{f_n(x)}{f(x)} &\geq \liminf_{n \rightarrow \infty} \frac{1}{(1 + \epsilon)f(y)} \inf_{x \in \Delta} f_n(x) \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{(1 + \epsilon)f(y)} \min_{j=0,1,\dots,d} \frac{P_n(\Delta_j)}{|\Delta|} \\ &= \frac{1}{(1 + \epsilon)f(y)} \min_{j=0,1,\dots,d} \frac{P(\Delta_j)}{|\Delta|} \geq \frac{1 - \epsilon}{1 + \epsilon} \end{aligned}$$

and

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{x \in \Delta} \frac{f_n(x)}{f(x)} &\leq \limsup_{n \rightarrow \infty} \frac{1}{(1 - \epsilon)f(y)} \sup_{x \in \Delta} f_n(x) \\ &\leq \frac{1}{(1 - \epsilon)f(y)} \left( \frac{P(\Delta)}{|\Delta|} \right)^{d+1} \left( \min_{j=0,1,\dots,d} \frac{P(\Delta_j)}{|\Delta|} \right)^{-d} \leq \left( \frac{1 + \epsilon}{1 - \epsilon} \right)^{d+1}. \end{aligned}$$

For  $\epsilon$  sufficiently small, both  $(1 - \epsilon)/(1 + \epsilon) \geq 1 - \eta$  and  $((1 + \epsilon)/(1 - \epsilon))^{d+1} \leq 1 + \eta$ , which proves the assertion of step 1.

**Step 2:** If  $f$  is continuous at  $y \in \mathbb{R}^d$  with  $f(y) = 0$ , then for any  $\eta > 0$  there exists a number  $\delta(y, \eta) > 0$  such that

$$\limsup_{n \rightarrow \infty} \sup_{x \in B(y, \delta(y, \eta))} f_n(x) \leq \eta.$$

For this step we employ Lemma 4.2. Let  $\delta_0 > 0$  such that  $B(0, \delta_0)$  is contained in  $\text{int}\{f > 0\}$ . Furthermore, let  $J_0 > 0$  be the minimum of  $f$  over  $B(0, \delta_0)$ . Then step 1 entails that

$$\liminf_{n \rightarrow \infty} \inf_{x \in B(0, \delta_0)} f_n(x) \geq J_0.$$

Moreover, for any  $t \in (0, 1)$  and  $\delta_t := (1 - t)\delta_0/(1 + t)$ ,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{x \in B(y, \delta_t)} f_n(x) &\leq J_0^{1-1/t} \limsup_{n \rightarrow \infty} \left( \frac{P_n(B(ty, \delta_t))}{|B(y, \delta_t)|} \right)^{1/t} \\ &\leq J_0^{1-1/t} \left( \frac{P(B(ty, \delta_t))}{|B(y, \delta_t)|} \right)^{1/t} \\ &\leq J_0^{1-1/t} \left( \sup_{x \in B(ty, \delta_t)} f(x) \right)^{1/t}. \end{aligned}$$



But the latter bound tends to zero as  $t \uparrow 1$ .

**Final step:**  $(f_n)_n$  converges to  $f$  uniformly on any closed set of continuity points of  $f$ .

Let  $S$  be such a closed set. Then Steps 1 and 2 entail that

$$\lim_{n \rightarrow \infty} \sup_{x \in S \cap B(0, \rho)} |f_n(x) - f(x)| = 0$$

for any fixed  $\rho \geq 0$ , because  $S \cap B(0, \rho)$  is compact, and any point  $y \in S \setminus \text{int}\{f > 0\}$  satisfies  $f(y) = 0$ .

On the other hand, let  $\Delta$  be a nondegenerate simplex with corners  $x_0, x_1, \dots, x_d \in \text{int}\{f > 0\}$ . Step 1 also implies that  $\lim_{n \rightarrow \infty} f_n(x_i) = f(x_i)$  for  $i = 0, 1, \dots, d$ , so that Lemma 3.3 entails that

$$\limsup_{n \rightarrow \infty} \sup_{x: \|x\| \geq \rho} \max\{f_n(x), f(x)\} \leq \max_{i=0, \dots, d} f(x_i) H\left(C \min_{i=0, \dots, d} f(x_i) (1 + \rho^2)^{1/2}\right) \quad (4.1)$$

for any  $\rho \geq 0$  with a constant  $C = C(x_0, \dots, x_d) > 0$ . Since this bound tends to zero as  $\rho \rightarrow \infty$ , the assertion of Theorem 2.1, Part (i) follows.  $\square$

Our proof of Theorem 2.1, Part (ii), is based on Part (i) and an elementary result about convex sets:

**Lemma 4.3.** *Let  $\mathcal{C}$  be a convex subset of  $\mathbb{R}^d$  containing  $B(0, \delta)$  for some  $\delta > 0$ . If  $y \in \mathcal{C}$ , then*

$$B(ty, (1-t)\delta) \subset \mathcal{C} \quad \text{for all } t \in [0, 1].$$

*If  $y \in \mathbb{R}^d \setminus \mathcal{C}$ , then*

$$B(\lambda y, (\lambda-1)\delta) \subset \mathbb{R}^d \setminus \mathcal{C} \quad \text{for all } \lambda \geq 1.$$

One consequence of this lemma is the well-known fact that the boundary of the convex set  $\{f > 0\}$  has Lebesgue measure zero. Namely, for any unit vector  $u \in \mathbb{R}^d$  there exists at most one number  $r > 0$  such that  $ru \in \partial\{f > 0\}$ . Lemma 4.3 is needed to obtain a refinement of this fact.

**Proof of Lemma 4.3.** By convexity of  $\mathcal{C}$  and  $B(0, \delta) \subset \mathcal{C}$ , it follows from  $y \in \mathcal{C}$  that

$$\mathcal{C} \supset \{(1-t)v + ty : v \in B(0, \delta)\} = B(ty, (1-t)\delta)$$

for any  $t \in [0, 1]$ . In case of  $y \notin \mathcal{C}$ , for  $\lambda \geq 1$  and arbitrary  $x \in B(\lambda y, (\lambda-1)\delta)$  we write  $x = \lambda y + (\lambda-1)v$  with  $v \in B(0, \delta)$ . But then

$$y = (1 - \lambda^{-1})(-v) + \lambda^{-1}x.$$

Hence  $y \notin \mathcal{C}$  is a convex combination of a point in  $B(0, \delta) \subset \mathcal{C}$  and  $x$ , so that  $x \notin \mathcal{C}$ , too.  $\square$

**Proof of Theorem 2.1, Part (ii).** It follows from (4.1) in the proof of Part (i) with  $\rho = 0$  that

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}^d} f_n(x) < \infty.$$

Since  $(f_n)_n$  converges to  $f$  pointwise on  $\mathbb{R}^d \setminus \partial\{f > 0\}$ , and since  $\partial\{f > 0\}$  has Lebesgue measure zero, dominated convergence yields

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d} \exp(A(x)) |f_n(x) - f(x)| dx \\ &= \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d \setminus B(0, \gamma)} \exp(A(x)) |f_n(x) - f(x)| dx \\ &\leq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d \setminus B(0, \gamma)} \exp(A(x)) \max(f_n(x), f(x)) dx \end{aligned}$$

for any fixed  $\gamma > 0$ .

It follows from Assumption (2.1) that for a suitable  $\rho > 0$ ,

$$A(x) + \varphi(x) - \varphi(0) \leq -1 \quad \text{whenever } \|x\| \geq \rho.$$

Utilizing sublinearity of  $A$  and concavity of  $\varphi$ , we may deduce that for  $x \in \mathbb{R}^d$  with  $\|x\| \geq \rho$  even

$$\begin{aligned} A(x) + \varphi(x) &= \varphi(0) + A(x) + \|x\| \frac{\varphi(\|x\|u) - \varphi(0)}{\|x\|} \\ &\leq \varphi(0) + A(x) + \|x\| \frac{\varphi(\rho u) - \varphi(0)}{\rho} \\ &= \varphi(0) + \rho^{-1} \|x\| (A(\rho u) + \varphi(\rho u) - \varphi(0)) \\ &\leq \varphi(0) - \rho^{-1} \|x\|, \end{aligned}$$

where  $u := \|x\|^{-1}x$ . In particular,  $\int_{\mathbb{R}^d} \exp(A(x)) f(x) dx$  is finite. Now let  $\delta > 0$  such that  $B(0, \delta) \subset \{f > 0\}$ . It follows from Lemma 4.3 that for any unit vector  $u \in \mathbb{R}^d$ , either  $2\rho u \in \{f > 0\}$  and  $B(\rho u, \delta/2) \subset \{f > 0\}$ , or  $2\rho u \in \{f = 0\}$  and  $B(3\rho u, \delta/2) \subset \{f = 0\}$ . Hence

$$K := \{0\} \cup \left\{ x \in \mathbb{R}^d : \|x\| \in \{\rho, 3\rho\}, \inf_{y \in \partial\{f > 0\}} \|x - y\| \geq \delta/2 \right\}$$

defines a compact subset of  $\mathbb{R}^d \setminus \partial\{f > 0\}$  such that

$$K \cap \{\rho u, 3\rho u\} \neq \emptyset \quad \text{for any unit vector } u \in \mathbb{R}^d.$$

According to Part (i),  $(f_n)_n$  converges to  $f$  uniformly on  $K$ . Thus for fixed numbers  $\epsilon' > 0$ ,  $\epsilon'' \in (0, \rho^{-1})$  and sufficiently large  $n$ , the log-densities  $\varphi_n := \log f_n$  satisfy the following inequalities:

$$\begin{aligned} A(ru) + \varphi_n(ru) &= \varphi_n(0) + r \left( A(u) + \frac{\varphi_n(ru) - \varphi_n(0)}{r} \right) \\ &\leq \varphi_n(0) + r \left( A(u) + \min_{s=\rho, 3\rho} \frac{\varphi_n(su) - \varphi_n(0)}{s} \right) \\ &\leq \varphi(0) + \epsilon' - \epsilon'' r \end{aligned}$$

for all unit vectors  $u \in \mathbb{R}^d$  and  $r \geq 3\rho$ . Hence for  $\gamma \geq 3\rho$ ,

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d \setminus B(0, \gamma)} \exp(A(x)) \max(f_n(x), f(x)) dx \\
& \leq f(0) \int_{\mathbb{R}^d \setminus B(0, \gamma)} \exp(\epsilon' - \epsilon'' \|x\|) dx \\
& = \text{const}(d) f(0) \int_{\gamma}^{\infty} r^{d-1} \exp(\epsilon' - \epsilon'' r) dr \\
& \rightarrow 0 \quad \text{as } \gamma \rightarrow \infty.
\end{aligned}$$

□

**Proof of Proposition 2.2.** It follows from convexity of  $\exp(\cdot)$  that  $\Theta(P)$  is a convex subset of  $\mathbb{R}^d$ , and obviously it contains 0. Now we verify it to be open. For any fixed  $\theta \in \Theta(P)$  we define a new probability density

$$\tilde{f}(x) := C^{-1} \exp(\theta^\top x) f(x) = \exp(\theta^\top x + \varphi(x) - \log C)$$

with  $C := \int_{\mathbb{R}^d} \exp(\theta^\top x) f(x) dx$ . Obviously,  $\tilde{f}$  is log-concave, too. Thus, by Corollary 3.4, there exist constants  $C_1, C_2 > 0$  such that  $\tilde{f}(x) \leq C_1 \exp(-C_2 \|x\|)$  for all  $x \in \mathbb{R}^d$ . In particular,

$$\infty > C \int_{\mathbb{R}^d} \exp(\delta^\top x) \tilde{f}(x) dx = \int_{\mathbb{R}^d} \exp((\theta + \delta)^\top x) f(x) dx$$

for all  $\delta \in \mathbb{R}^d$  with  $\|\delta\| < C_2$ . This shows that  $\Theta(P)$  is open.

Finally, let  $\theta \in \Theta(P)$  and  $\epsilon > 0$  such that  $B(\theta, \epsilon) \subset \Theta(P)$ . With the previous arguments one can show that for each unit vector  $u \in \mathbb{R}^d$  there exist constants  $D(u) \in \mathbb{R}$  and  $C(u) > 0$  such that  $(\theta + \epsilon u)^\top x + \varphi(x) \leq D(u) - C(u) \|x\|$  for all  $x \in \mathbb{R}^d$ . By compactness, there exist finitely many unit vectors  $u_1, u_2, \dots, u_m$  such that the corresponding closed balls  $B(u_i, (2\epsilon)^{-1} C(u_i))$  cover the whole unit sphere in  $\mathbb{R}^d$ . Consequently, for any  $x \in \mathbb{R}^d \setminus \{0\}$  and its direction  $u(x) := \|x\|^{-1}x$ , there exists an index  $j = j(x) \in \{1, \dots, m\}$  such that  $\|u(x) - u_j\| \leq (2\epsilon)^{-1} C(u_j)$ , whence

$$\begin{aligned}
\theta^\top x + \epsilon \|x\| + \varphi(x) &= (\theta + \epsilon u(x))^\top x + \varphi(x) \\
&\leq (\theta + \epsilon u_j)^\top x + \varphi(x) + \epsilon \|u_j - u(x)\| \|x\| \\
&\leq D(u_j) + (\epsilon \|u_j - u(x)\| - C(u_j)) \|x\| \\
&\leq \max_{i=1, \dots, m} D(u_i) - 2^{-1} \min_{i=1, \dots, m} C(u_i) \|x\| \\
&\rightarrow -\infty \quad \text{as } \|x\| \rightarrow \infty.
\end{aligned}$$

□

**Proof of Theorem 2.3.** As mentioned already, the statements about  $\theta \in \Theta(P)$  and  $\Pi(\cdot)$  are a consequence of Theorem 2.1 (ii) and Proposition 2.2. Note also that for  $\theta \in \mathbb{R}^d \setminus \Theta(P)$  and

arbitrary  $r > 0$ ,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d} \exp(\theta^\top x) P_n(dx) &\geq \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \min(\exp(\theta^\top x), r) P_n(dx) \\ &= \int_{\mathbb{R}^d} \min(\exp(\theta^\top x), r) P(dx), \end{aligned}$$

and the right hand side tends to infinity as  $r \uparrow \infty$ .  $\square$

## References

- [1] M. AN (1998). Log-concavity versus log-convexity. *J. Econometric Theory* **80**, 350–369.
- [2] M. BAGNOLI and T. BERGSTROM (2005). Log-concave probability and its applications. *Econometric Theory* **26**, 445–469.
- [3] R. R. BAHADUR and L. J. SAVAGE (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27**, 1115–1122.
- [4] M. L. CULE and L. DÜMBGEN (2008). On an auxiliary function for log-density estimation. Technical report 71, IMSV, University of Bern. (arXiv:0807.4719)
- [5] M. L. CULE, R. B. GRAMACY and R. J. SAMWORTH (2009). LogConcDEAD: An R package for maximum likelihood estimation of a multivariate log-concave density. *Journal of Statistical Software* **29**(2).
- [6] M. L. CULE and R. J. SAMWORTH (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4**, 254–270.
- [7] M. L. CULE, R. J. SAMWORTH and M. I. STEWART (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. R. Statist. Soc. B (with discussion)*, to appear. (arXiv:0804.3989)
- [8] S. DHARMADHIKARI and K. JOAG-DEV (1988). *Unimodality, Convexity, and Applications*. Academic Press, London.
- [9] D. L. DONOHO (1988). One-sided inference about functionals of a density. *Ann. Statist.* **16**, 1390–1420.
- [10] L. DÜMBGEN and K. RUFIBACH (2009). Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. *Bernoulli* **15**(1), 40–68.

- [11] L. DÜMBGEN and K. RUFIBACH (2011). logcondens: Computations related to univariate log-concave density estimation. *J. Statist. Software* **39**(6).
- [12] L. DÜMBGEN, A. HÜSLER and K. RUFIBACH (2007). Active set and EM algorithms for log-concave densities based on complete and censored data. Technical report 61, IMSV, University of Bern. (arXiv:0707.4643)
- [13] A. HÜSLER (2008). *New aspects of statistical modeling with log-concave densities*. Ph.D. thesis, IMSV, University of Bern.
- [14] R. LANG (1986). A note on the measurability of convex sets. *Arch. Math.* **47**(1), 90–92.
- [15] P. MASSART (1990). The tight constant in the Dvoretzki-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18**, 1269–1283.
- [16] J. PAL, M. WOODROOFE and M. MEYER (2006). Estimating a Polya frequency function. In: *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond* (R. Liu, W. Strawderman, C.-H. Zhang, eds.) , *IMS Lecture Notes and Monograph Series* **74**, 239–249. Institute of Mathematical Statistics.
- [17] D. SCHUHMACHER and L. DÜMBGEN (2010). Consistency of multivariate log-concave density estimators. *Statist. Probab. Lett.* **80**(5-6), 376–380.
- [18] A. W. VAN DER VAART and J. A. WELLNER (1996). *Weak Convergence and Empirical Processes, with Applications to Statistics*. Springer Series in Statistics. Springer-Verlag, New York.
- [19] G. WALTHER (2009). Inference and modeling with log-concave distributions. *Statist. Sci.* **24**(3), 319–327.